Impact of model uncertainty on sensitivity analysis

*Julien JACQUES *, Christian LAVERGNE *, Nicolas DEVICTOR ***

* INRIA Rhône-Alpes
655, avenue de l'Europe - Montbonnot Saint-Martin
38334 Saint-Ismier Cedex FRANCE
phone (+33) 476 615 342 fax (+33) 476 615 477
Email : julien.jacques@inrialpes.fr

** CEA Cadarache
DER/DEN/STR/LCFR Bat 208
13108 St Paul lez Durance Cedex

1 Introduction

In fields of numerical modelling, sensitivity analysis is often used for model calibration or model validation, and to find which variables mostly contribute to output variability. In this context, we propose to study the impact of model uncertainty on results of sensitivity analysis. We will focus in this abstract on what means model uncertainty for us, and how we propose to treat it. Finally, we will present applied motivations, in the domain of nuclear safety.

2 Context of sensitivity analysis

The name of sensitivity analysis is currently used, for different means. Here, we will consider, like named in [3], global sensitivity analysis techniques, based on the study of the variances of model variables. Those methods consist in the computation of sensitivity indices, which quantify the sensitivity of model output variance to model inputs. For a model

$$Y = f(X_1, ..., X_n),$$

first order sensitivity indices are defined by

$$S_i = \frac{V(E[Y|X_i])}{V(Y)},$$

and express the part of variance of model output Y due to model input X_i. Higher order indices are also defined, to express effect of input interaction and total indices for total effect of one input. Methods of estimation of those indices are introduced by Cukier (FAST [1], [4]), Sobol [5], McKay [2], among others. We will use Sobol method for numerical experiments.

3 Model uncertainty and aim of this work

In many fields like reliability of mechanical structures, behavior of thermohydraulic systems, or nuclear safety, mathematical models are used, at the same time for simulation, when experiments are too expensive or even impracticable (nuclear accident), and for prediction. We will consider model uncertainty in two particular

ways.
Firstly, we consider uncertainty due to the use of a simplified model M_{simp}, while a reference model M_{ref} exists. This is useful when computational time for M_{ref} run is too long.
Secondly, we consider uncertainty like a mutation of the studied process, modelled by a start model M_{start}. This mutation can be due to new information or changes in the process. Thus, we get a new process modelled by a new model M_{mut}.
The aim of this work is to take into account those uncertainties in sensitivity analysis. When M_{simp} is a simplification of M_{ref}, is it possible to give more precise results on M_{ref} sensitivity indices, that doing the approximation by M_{simp} ones ? And when M_{mut} is a mutation of M_{start}, can we obtain sensitivity results on M_{mut} from M_{start} ones ?

4 Methodology

We present here the methodology employed to treat the two ways of seeing model uncertainty.

Uncertainty due to the use of simplified model
As introduced previously, assume that the working model is

$$M : Y = f(X_1, ..., X_n),$$

where $(X_1, ..., X_n)$ and Y are respectively input and output variables, and f is a mathematical function with long computation time (maybe several hours by model run). So, it's impossible to estimate sensitivity indices with Sobol method, because estimation needs too much model runs and also too much time. We find the same problem with other estimation methods.
Assume now that we have been able to do N model simulations, in other words, for N simulations of the inputs, we have compute the N corresponding outputs. In practice, N can be almost equal to some hundreds, or maybe one thousand.
As these N simulations are insufficient to estimate directly sensitivity indices, we adjust a response surface on the model, with these N simulations. This response surface is considered like a simplified model, that we named

$$M_{simp} : Y_{simp} = f_{simp}(X_1, ..., X_p),$$

with $p \leq n$. One particularity of this response surface is that its computation time is very quick, and hence we can do as many simulations as we want.
So, it's possible to approximate M sensitivity indices by M_{simp} ones, but quality of this estimation will depend from quality of response surface adjustment.
For improve quality of this estimation, we will take into account residuals ϵ obtained during adjustment :

$$Y = Y_{simp} + \epsilon.$$

One aim of this work is to succeed to characterize ϵ.
We propose two solutions. The first method is empirical : it consists in adjustment of a response surface on residuals $(\epsilon^i)_{i=1..N}$. Envisaged response surfaces are polynomial functions, linear or additive models. The second method is more analytical, and consists to define a distance between f and f_{simp}, the two functions of models M and M_{simp}.
Those two methods of space between models characterization consist in definition and estimation of

$$\Delta \simeq Y - Y_{simp},$$

and allow us to obtain a better estimation of sensitivity indices, in introducing correction function to Δ, or to obtain information on confidence on this estimation (for example with confidence intervals).

Uncertainty due to process mutation
The second way of considering uncertainty, is to envisage a process mutation, translated in model mutation. We don't wish to compute sensitivity indices once again. The method we investigate consists in drawing up a typology of possible model mutations, and to define, according to the identified mutation, how to estimate new sensitivity indices with minimal cost. Mutations can be addition or deletion of variables, affine transformation of the models, and so on.
For example, consider the model :

$$M : Y = f(X_1)g(X_2, ..., X_n),$$

and assume that the mutation consists in fixing X_1 to a single value α. Hence, M mute in M_{mut} defined by :

$$M_{mut} : Y_{mut} = f(\alpha)g(X_2, X_3).$$

We prove that M_{mut} sensitivity indices can be estimate with M ones, by :

$$S_j^{mut} = (\frac{f(\alpha)}{f(E[X_i])})^2 S_j \times \frac{V(Y)}{V(Y_{mut})},$$

for $2 \leq j \leq n$. There exists similar formulas for total and higher order sensitivity indices.
Hence, we can obtain M_{mut} sensitivity indices unless to do a new complete analysis.

5 Application in nuclear safety

Nuclear central chronical rejects : consequences for people
We have a computer code which models chronic atmospherical rejects, quantifies environmental transferts, and mesures impact on people, for two nuclear facilities. Response surfaces have been adjusted on this model, because its computational time is too long to be usable. Hence, sensitivity indices have been estimated, to know which variables are the most important, and thus which variables must be better known, to low down prediction uncertainties.
The first application made on this study, is to take into account the use of response surfaces in sensitivity indices estimation, while reference model exists. Second application is to adapt sensitivity results to another nuclear facility, by modifying geographical and environmental model parameters, and thus by introducing a model mutation.

References

[1] R.I.Cukier, H.B.Levine, K.E.Shuler. *Nonlinear Sensitivity Analysis of Multiparameter Model Systems.* Journal of Computational Physics, 1978.

[2] M.D.McKay. *Evaluating Prediction uncertainty. Technical Report NUREG/CR-6311.* US Nuclear Regulatory Commission and Los Alamos National Laboratory, 1995.

[3] A.Saltelli, K.Chan, E.M.Scott. *Sensitivity Analysis.* Wiley, 2000.

[4] A.Saltelli, S.Tarantola, K.Chan. *A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output.* Technometrics, Vol.41, No.1, 1999.

[5] I.M.Sobol. *Sensitivity Estimates for Nonlinear Mathematical Models.* Mathematical Modelling and Computational Experiments, 1993.